

University of Groningen

Statistics and EURO2004 (updated 24 June 2004)

Koning, Ruud H.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2004

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Koning, R. H. (2004). *Statistics and EURO2004 (updated 24 June 2004)*.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Statistics and EURO2004

Ruud H. Koning

Update 24th June 2004

Updated winning probabilities after the first round in the group stage can be found in the section 5. New updates after the groups stage are presented in section 6.

1 INTRODUCTION

Who will win the European Cup in Portugal? Many people have an opinion on this issue, and of course it would be profitable if one could look in a crystal ball and know the result beforehand. Some people expect statisticians to have such a crystal ball, and fortunately, that is not the case. Statisticians cannot predict the result of soccer matches with certainty, but they do have methods to estimate the probabilities of different outcomes (see for example Maher (1982), Clarke and Norman (1995), and Dixon and Coles (1997)). In this note I use the methodology of the Dutch *Werkgroep Voetbal & Statistiek* which is described detail in Koning, Koolhaas, Renes, and Ridder (2003) to estimate the winning probabilities of the participants of EURO2004. Our model consists of two parts. The first part estimates scoring intensities, that is, the number of goals one country is expected to score against another country. The second part is a simulation/probability model where these estimates of the the strengths of the participants is used to to identify which team is more likely to win the tournament than other teams.

2 ESTIMATION OF SCORING INTENSITIES

We estimate the relative strength of the participants using match-specific scoring intensities. In this approach, the expected number of goals by team i against team j , λ_{ij} , varies with team j . Team i is expected to score more if team j is a weak team than if team j is a strong team. Consider a match between Germany and England. During the period covered by the dataset, England conceded on average 0.85 goals per match. Germany played Scotland during the qualification tournament for EURO2000, and won by 2-1. During the sample period, Scotland concedes on average 0.89 goals per match. Based on the number of goals scored by Germany against Scotland, we expect Germany to score $2 \times \frac{0.85}{0.89} = 1.91$ against England. We expect Germany to score fewer goals against England because their defense is better than the defense from Scotland. Of course, Germany played more opponents than only Scotland, so we average over all opponents played by Germany.

Based on this approach of estimating scoring intensities by taking weighted averages, we can calculate the probabilities of the result of a game. The number of goals by team i against team j is assumed to follow a Poisson distribution. Tables 1 to 4 give information on the winning probabilities for the games in the first stage of the tournament. The entries should be read as follows: Netherlands-Germany 0.49/0.27/0.25

Table 1: Probability distribution outcomes 1

Group 1	Portugal	Spain	Greece	Russia
Portugal		0.27/0.23/0.49	0.66/0.19/0.14	0.39/0.25/0.35
Spain	0.49/0.23/0.27		0.73/0.18/0.094	0.49/0.26/0.25
Greece	0.14/0.19/0.66	0.094/0.18/0.73		0.16/0.24/0.6
Russia	0.35/0.25/0.39	0.25/0.26/0.49	0.6/0.24/0.16	

Table 2: Probability distribution outcomes 2

Group 2	England	France	Croatia	Switzerland
England		0.17/0.23/0.61	0.25/0.37/0.37	0.53/0.24/0.23
France	0.61/0.23/0.17		0.44/0.35/0.21	0.79/0.14/0.077
Croatia	0.37/0.37/0.25	0.21/0.35/0.44		0.53/0.30/0.17
Switzerland	0.23/0.24/0.53	0.077/0.14/0.79	0.17/0.30/0.53	

Table 3: Probability distribution outcomes 3

Group 3	Sweden	Bulgaria	Italy	Denmark
Sweden		0.36/0.29/0.35	0.23/0.28/0.49	0.51/0.24/0.25
Bulgaria	0.35/0.29/0.36		0.22/0.31/0.47	0.48/0.27/0.25
Italy	0.49/0.28/0.23	0.47/0.31/0.22		0.63/0.22/0.15
Denmark	0.25/0.24/0.51	0.25/0.27/0.48	0.15/0.22/0.63	

Table 4: Probability distribution outcomes 4

Group 4	Netherlands	Germany	Czech Republic	Latvia
Netherlands		0.49/0.27/0.25	0.42/0.27/0.31	0.77/0.15/0.08
Germany	0.25/0.27/0.49		0.29/0.28/0.42	0.63/0.22/0.15
Czech Republic	0.31/0.27/0.42	0.42/0.28/0.29		0.7/0.19/0.11
Latvia	0.08/0.15/0.77	0.15/0.22/0.63	0.11/0.19/0.7	

(table 4) means that The Netherlands wins the game against Germany with probability 0.49, and loses with probability 0.25. The game ends in a draw with probability 0.27. The Netherlands is favorite to win this game. The entry for the game Germany-Netherlands is of course the mirror image: 0.25/0.27/0.49.

3 RESULTS OF THE SIMULATION/PROBABILITY MODEL

Which countries are most likely to proceed to the second stage of the tournament? In tables 5 to 8 we see the probabilities for each country on each position. Focusing

Table 5: Probability distribution ranking group 1

Group 1	1	2	3	4
Portugal	0.2764	0.3207	0.2771	0.1258
Spain	0.4615	0.2935	0.1885	0.0565
Greece	0.0314	0.0895	0.2310	0.6481
Russia	0.2307	0.2963	0.3034	0.1696

Table 6: Probability distribution ranking group 2

Group 2	1	2	3	4
England	0.1560	0.2818	0.3401	0.2221
France	0.5592	0.2669	0.1328	0.0411
Croatia	0.2427	0.3327	0.2831	0.1415
Switzerland	0.0421	0.1186	0.2440	0.5953

Table 7: Probability distribution ranking group 3

Group 3	1	2	3	4
Sweden	0.2313	0.2912	0.2706	0.2069
Bulgaria	0.2124	0.2758	0.2911	0.2207
Italy	0.4702	0.2712	0.1752	0.0834
Denmark	0.0861	0.1618	0.2631	0.4890

Table 8: Probability distribution ranking group 4

Group 4	1	2	3	4
Netherlands	0.4559	0.3090	0.1813	0.0538
Germany	0.2087	0.2978	0.3552	0.1383
Czech Republic	0.3089	0.3256	0.2752	0.0903
Latvia	0.0265	0.0676	0.1883	0.7176

on group 4 (table 8), we see that the probability that The Netherlands wins group 4 is almost 46%. The probability that The Netherlands end either of the first or second place is 76%.

Conditionally on the results of the first stage, it is known which countries can meet during the remainder of the tournament. Using the same methodology to estimate scoring intensities, it is possible to calculate the winning probability for each team that survives the first round. The results are given in table 9. We see that France is the team most likely to win the title: their winning probability is 29.2%. They are followed by Italy and Spain. The Netherlands is the fourth most likely team to win the title. Note the dismal performance of the hosts: Portugal has a winning probability of 6.3%, even

Table 9: Success probabilities EURO2004

Country	Probability of			
	Quarter final	Semi final	Final	Champion
France	0.83480	0.60304	0.42542	0.29220
Italy	0.73460	0.46759	0.25626	0.14222
Spain	0.73240	0.36148	0.20415	0.10274
Netherlands	0.77860	0.43095	0.20167	0.09747
Croatia	0.56640	0.30548	0.16330	0.07647
Portugal	0.66420	0.28774	0.14660	0.06347
Czech Republic	0.62700	0.30514	0.12704	0.05307
Sweden	0.51700	0.24849	0.09680	0.03756
Bulgaria	0.49120	0.23707	0.09391	0.03698
England	0.43760	0.19862	0.09062	0.03375
Russia	0.49940	0.18261	0.07977	0.02932
Germany	0.50460	0.20885	0.07351	0.02519
Denmark	0.25720	0.08743	0.02254	0.00589
Switzerland	0.16120	0.04410	0.01286	0.00285
Greece	0.10400	0.01693	0.00360	0.00059
Latvia	0.08980	0.01447	0.00195	0.00023

though they have the home advantage. The reason for this is simple: Portugal is likely to end second in group A, and France is expected to win group B. Hence, Portugal and France are likely to meet in the quarter final, which France is likely to win, considering its recent record.

The probabilities in the last column should be compared to 6.25%, the probability that a team would win if they were all of equal quality. France is 4.5 times as likely to win than under this equal quality assumption.

The scoring intensities can also be used to simulate a full competition. These results are in table 10. France has a very high probability to win such a full competition. This is expected, since in a tournament one bad game can result in elimination while in a competition there are more games to show the quality of a team.

4 CONCLUSIONS

France is the most likely winner of Euro 2004, with Italy a distant second. The probability that France retains its title is almost 30%. The probability that the Netherlands wins its second European cup is almost 10%, which is still larger than the probability of winning the title if all teams were of equal quality.

During the course of the tournament, more information on the relative strengths of the participants becomes available, and the uncertainty of the outcome of games gets resolved. Updates to this document will be published at www.soccer-research.com at irregular intervals.

Table 10: Full competition simulation EURO2004

Country	Probability of			
	1st place	2nd place	3rd place	4th place
France	0.5239	0.2012	0.1041	0.0632
Italy	0.1270	0.1755	0.1508	0.1219
Spain	0.1061	0.1507	0.1391	0.1270
Portugal	0.0642	0.1047	0.1138	0.1114
Netherlands	0.0601	0.1065	0.1118	0.1140
Croatia	0.0289	0.0674	0.0908	0.0981
Czech Republic	0.0219	0.0459	0.0598	0.0720
England	0.0195	0.0357	0.0575	0.0690
Sweden	0.0168	0.0373	0.0491	0.0668
Bulgaria	0.0147	0.0295	0.0495	0.0598
Russia	0.0119	0.0284	0.0418	0.0493
Germany	0.0042	0.0128	0.0227	0.0310
Denmark	0.0006	0.0028	0.0057	0.0108
Switzerland	0.0002	0.0014	0.0033	0.0050
Greece	0.0000	0.0002	0.0002	0.0007
Latvia	0.0000	0.0000	0.0000	0.0000

Table 11: Success probabilities EURO2004 (16 June 2004)

Country	Probability of			
	Quarter final	Semi final	Final	Champion
France	0.962	0.740	0.522	0.354
Spain	0.918	0.494	0.286	0.138
Italy	0.654	0.408	0.216	0.119
Netherlands	0.722	0.402	0.185	0.087
Croatia	0.535	0.292	0.157	0.071
Czech Republic	0.728	0.362	0.145	0.060
Sweden	0.816	0.391	0.153	0.056
Germany	0.515	0.220	0.077	0.025
Portugal	0.333	0.112	0.057	0.025
England	0.321	0.143	0.064	0.021
Bulgaria	0.213	0.102	0.039	0.016
Russia	0.281	0.085	0.037	0.014
Denmark	0.317	0.108	0.027	0.007
Switzerland	0.182	0.057	0.017	0.004
Greece	0.469	0.077	0.017	0.003
Latvia	0.035	0.006	0.001	0.000

5 UPDATE 16 JUNE 2004

During the tournament, more information becomes available. First, we get a better impression of the strengths of the participating teams. This leads to new estimates of the scoring intensities, which are based on more games. Second, the uncertainty resolves itself during the tournament. The expected ranking in a group is after one round of games differs from the one before the start of the tournament.

Using the additional information of the eight games played so far, we recalculated the winning probabilities. They are listed in table 11. The result of France against England has improved France's probability to reach the quarter final. The probability that the hosts win the title has decreased from 6.3% to 2.5% as a result of their unexpected loss against Greece.

6 UPDATE 23 JUNE 2004

The group stage is finished and the games of the quarter finals are known. Table 11 lists the probabilities of the remaining teams to reach the semi finals and the title. France is still the favorite for the title, mainly because the model gives a very high probability of reaching the semi final. Also, note that the Czech Republic has a quite high probability of reaching the semi finals. Portugal is slightly favorite to win against England, and Holland and Sweden are in for a tight game.

CONTACT

Ruud Koning (www.rhkoning.com) is professor of sports economics at the University of Groningen, and does research on the intersection of economics, statistics, and sports. He can be reached at ruud@rhkoning.com, or by telephone at +31 (0)50 363 7192 or

Table 12: Success probabilities EURO2004 (20 June 2004)

Country	Probability of			
	Quarter final	Semi final	Final	Champion
France	1	0.905	0.695	0.498
Portugal	1	0.563	0.288	0.110
Netherlands	1	0.514	0.268	0.104
Czech Republic	1	0.650	0.207	0.099
Sweden	1	0.486	0.246	0.093
England	1	0.437	0.198	0.068
Denmark	1	0.350	0.069	0.022
Greece	1	0.095	0.028	0.006
Spain	0	0.000	0.000	0.000
Russia	0	0.000	0.000	0.000
Croatia	0	0.000	0.000	0.000
Switzerland	0	0.000	0.000	0.000
Bulgaria	0	0.000	0.000	0.000
Italy	0	0.000	0.000	0.000
Germany	0	0.000	0.000	0.000
Latvia	0	0.000	0.000	0.000

Table 13: Success probabilities EURO2004 (20 June 2004)

Country	Probability of			
	Quarter final	Semi final	Final	Champion
France	1	0.905	0.695	0.498
Portugal	1	0.563	0.288	0.110
Netherlands	1	0.514	0.268	0.104
Czech Republic	1	0.650	0.207	0.099
Sweden	1	0.486	0.246	0.093
England	1	0.437	0.198	0.068
Denmark	1	0.350	0.069	0.022
Greece	1	0.095	0.028	0.006
Spain	0	0.000	0.000	0.000
Russia	0	0.000	0.000	0.000
Croatia	0	0.000	0.000	0.000
Switzerland	0	0.000	0.000	0.000
Bulgaria	0	0.000	0.000	0.000
Italy	0	0.000	0.000	0.000
Germany	0	0.000	0.000	0.000
Latvia	0	0.000	0.000	0.000

+31 (0)62 224 6443.

REFERENCES

- Clarke, S.R. and J.M. Norman (1995). Home ground advantage of individual clubs in English soccer. *The Statistician* 44(4), 509–521.
- Dixon, M.J. and S.G. Coles (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* 46, 265–280.
- Koning, R.H., M. Koolhaas, G. Renes, and G. Ridder (2003). A simulation model for football championships. *European Journal of Operational Research* 148(2), 268–276.
- Maher, M.J. (1982). Modelling association football scores. *Statistica Neerlandica* 36, 109–118.